View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Munich RePEc Personal Archive

MPRA
Munich Personal RePEc Archive

The effects of body-worn cameras on police efficiency: A study of local police agencies in the US.

Alda, Erik

American University

31 October 2020

Online at https://mpra.ub.uni-muenchen.de/103994/
MPRA Paper No. 103994, posted 26 Nov 2020 13:04 UTC

The effects of body-worn cameras on police efficiency: A study of local police agencies in the US

Erik Alda, PhD*

PRELIMINARY DRAFT-COMMENTS WELCOME

First draft: October 31, 2020; This draft: November 24, 2020

Abstract

Do Body-Worn Cameras improve police efficiency? This study answers this question in the context of a sample of local police agencies in the US, where the adoption of BWCs by police agencies has increased significantly in recent years. To estimate the effects of BWCs on police efficiency, this study exploited the differences in the adoption of BWCs between agencies that acquired them ("acquirers") and agencies that deployed them ("deployers"). Using a multiple stage approach, in the first stage the author estimated the efficiency of local police agencies using a robust order-m model. In the second stage, the author estimated the effects of BWCs using a range of matching estimators and an instrumental variable model. The first stage results show that police agencies could improve their efficiency by 31 percent from 0.76 to 1. The second stage matching and IV estimates suggest that BWCs can help improve police efficiency between eight and 21 percentage points. The effects are larger for those agencies that fully deployed BWCs with their officers. Overall, this study's results support the argument that BWCs can help improve police efficiency.

Keywords: Police, Performance, Efficiency, Data Envelopment Analysis, Matching Estimators, Instrumental Variables.

*Faculty Fellow- Department of Justice, Law and Criminology, American University.

1 Introduction

During the last five years, there has been a dramatic increase in the use of Body-Worn Cameras (BWCs)across law enforcement agencies in the US. Police departments of all sizes have acquired and deployed BWCs to improve their transparency, accountability, and performance (Chapman, 2018). Simultaneously, research on the impact of BWCs on a wide range of law enforcement outcomes has also burgeoned.

A growing body of empirical evidence provides support for the use of this technology to improve various police outcomes, such as accountability, reductions of civilian complaints against police, police-citizen interactions, citizen behavior, among others (Lum, Stoltz, Koper,& Scherer, 2019). However, research on the effects of BWCs on police efficiency, however, remains unexamined. This study addresses some of these questions by estimating the effects of BWCs on the efficiency of local police agencies.

To answer this question, the author first estimated police agencies' efficiency using well-known methods to measure efficiency in organizations such as Data Envelopment Analysis (Charnes & Cooper, 1957). In particular, the author employed a robust approach–order-*m*– (Cazals & Florens, 2007) that corrects for known biases in efficiency measurement, such as the presence of outliers and measurement error.

Secondly, the author used a range of matching methods and instrumental variable regression to assess the effect of BWCs on police efficiency between agencies. The use of BWCs by police agencies varies widely. The data show that about 60%[1] of agencies acquired BWCs compared to those that did not. However, not all of the agencies deployed BWCs with their officers. In fact, out of the 60% of agencies that acquired BWCs, 84% implemented a partial or full deployment; and only 40% of the 84% of agencies that deployed BWCs implemented a full deployment with their officers.

The author exploited this difference between BWCs "acquirers" and BWCs

[1] This percentage is based on the final sample used for the analyses, which was 615 local police agencies. See the section on data for more information.

”deployers” and conceptualized the ”acquirers” as Intent to Treat (ITT) and the ”deployers” as Treatment on the Treated (TOT). This difference in the adoption of BWCs allowed me to match agencies on a set of organizational and environmental characteristics and assess differences in efficiency levels between ”acquirers” and ”non-acquirers”. Then, the author used the ”acquirers” (ITT) measure as an instrumental variable to examine differences in efficiency between ”deployers[2]” and ”acquirers” using LATE[3] analyses.

The findings indicated that BWCs increase police efficiency between seven and 12 percentage points for the ITT analyses and between 10 and 21 percentage points for the LATE estimates, thus supporting arguments that this technology contributes to improving police efficiency and increasing transparency and accountability in police organizations.

This study’s results contribute to the rapidly growing literature on the use of BWCs in various ways. First, to my knowledge, this is the first study that examines the effect of BWCs on police efficiency. The current scholarly and policy literature on this topic focuses mainly on measuring the effects of BWCs on outcomes such as transparency, accountability, legitimacy, and on criminal resolution, intelligence gathering, and criminal justice processes outcomes. Some studies have examined outcomes like the speed of criminal resolution or criminal justice processes outcomes. These outcomes approximate an efficiency measure since criminal investigations, for example, are critical processes of a police production function because they may lead to more crimes cleared (c.f. Morrow, Katz, & Choate, 2016; Owens, Mann, & McKenna, 2014). However, none of these studies use efficiency as their primary focus of research, nor do they produce an actual efficiency estimate. Hence, in addition to examining the effect that BWCs have on police efficiency, the author borrowed from the literature on productive efficiency and provide an estimate of the levels of police efficiency by using a range and inputs and how their combination contributes to police output (Charnes, Cooper, & Rhodes, 1978).

Second, this study focuses on a sample of 615 police agencies instead of, for example, a single agency or a subset of agencies within a police district where most

[2]These are the agencies that are assumed compliant and deploy the BWCs.

[3]LATE stands for Local Average Treatment Effects. It is the same as Treatment on the Treated (TOT) effects.

studies draw their experimental or quasi-experimental evidence from (Kim, 2019; Ariel et al.,2016; Jennings et al.,2017; Harcourt & Ludwig,2006). Although the strength and robustness of results from well-designed experiments are irrefutable, the results of this study are useful because they reveal effects across a larger number of police agencies and help support the results found in experimental and quasi-experimental evaluations.

Third, the results of this study can offer useful operational insights for police agencies becuause the deployment of BWCs can assist them in having higher clearance rates because of the faster availability of critical information to help them resolve crimes. In turn, efficiency gains resulting from BWCs can help strengthen other important areas of police operations.

The remainder of the study is structured as follows. Section 2 presents a review of the literature on the use of BWCs, which has focused mainly on experimental evidence assessing BWCs's efficacy on a broad range of outcomes related to officer and citizen behavior, police use of force, civilian complaints, and police accountability, among others. Section 3 presents the data used in the analyses. Section 4 presents and discusses the multiple stage empirical approach to first estimate the efficiency scores and then examine the effects of BWCs using efficiency as the primary outcome of interest. Section 5 presents the results of the preferred model and several robustness tests and sensitivity of the results to the presence of hidden bias. Finally, Section 6 presents the conclusions and discusses the study's limitations.

2 Literature Review

There has been rapid growth in the literature on the adoption of this technological innovation in law enforcement and its resulting impacts on a wide range of outcomes in the last five years (Lum, Koper, Merola, Scherer, & Reioux, 2015). Scholars have categorized research on the impact of BWCs around six main areas of study, including impacts on officer behavior and citizen behavior; officer attitudes about BWCs; citizen and community attitudes about police or cameras; criminal investigations; and police organizational structure (Lum, Stoltz, Koper, &

Scherer, 2019).

The evidence around the impacts of BWCs police efficiency remains largely understudied. Studies on the effects of BWCs on criminal investigations and crime resolution are, however, closest to efficiency measurement. Crime investigations are a critical component of a police production function and are often used as a measure of police organizations' efficiency. For example, the time it takes to clear crimes and the number of resources saved from using BWCs could be interpreted as a measure of efficiency, and, in fact, previous research on police efficiency has used these variables as outputs in an efficiency model (c.f. Alda, 2014; Alda & Dammert, 2019). Thus, this literature review focuses on the strand of research that more closely approximates the study of efficiency as an outcome, although no studies to date have used a measure of police efficiency as their primary outcome of interest. For a thorough review of available evaluations and research on BWCs, see Lum et al. (2019).

The number of research studies focusing on this proxy of police efficiency, however, is relatively small, accounting for just 6% of all the published research on BWCs to date (Lum et al., 2019), and the results are mixed. Studies have examined the impact of BWCs using the gold standard for evaluations (RCTs) or quasi-experimental approaches, and "before and after" approaches as well as qualitative analyses to support their quantitative findings.

Yokum, Ravishankar, and Coppock (2017) conducted an RCT with more than 2,000 police officers in Washington DC's MPD to examine the impact of BWCs on police complaints, police use of force, policing activity, and judicial outcomes. It is noteworthy that the latter approximates a measure of efficiency in that it captures the process in which police arrests are prosecuted in the justice system because the footage produced by BWCs can lead to faster case resolution (Yokum et al., 2017). Overall, the study found very small effects, none of which were statistically significant. One potential explanation for the lack of results is that the researchers did not have access to the full prosecutorial datasets but only dataset with the police department's initial charges.

While Yokum et al. (2017) offer a range of thorough explanations for the lack of results, the simpler and most likely explanation is that BWCs do not affect the outcomes studied. In the case of the efficiency proxy, the camera footage did not affect judicial outcomes. The study concludes by nuancing the message around the expectations of BWCs as well as encouraging more research on the impact of BWCs (Yokum et al., 2017).

Owens, Mann, and McKenna (2014) also conducted an RCT to measure the impact of BWCs using sample of 308 police officers in Essex, UK, focusing on reducing bias in the results of incidents attended by officers. The authors also interviewed officers in the treatment group to better understand the operational challenges of BWC deployment.

Their findings suggest no differences in the number of incidents sanctioned between officers who wore BWCs and those who did not. However, they suggest significant differences in the type of detected sanction that resulted in criminal charges in the treatment group compared to the control group–81% vs. 72%, respectively.

The study's qualitative findings indicated that officers who used BWCs experienced more accountability and paid more attention to their behavior while conducting policing activities. The study concludes with a hypothetical statement that BWCs could be useful in increasing the proportion of detected offenses resulting in criminal charges, particularly in domestic abuse cases (Owens et al., 2014).

A recent study used the LEMAS BWCs supplement survey to examine the causal impact of BWCs on a range of performance and police use of force outcomes by exploiting the variation in the adoption of BWCs adoption (Kim, 2019). Kim's study departs from previous research on BWCs because it examines the impact using a national survey of over 1,000 agencies instead of a single agency or group of agencies within a police district. The princial finding suggests a 54% drop in citizen deaths resulting from police use of force. Furthermore, the study argued that investing in BWCs could yield substantial benefits to police agencies in reducing lawsuits resulting from use of force incidents (Kim, 2019).

Katz and colleagues (2014) and Morrow, Katz, and Choate (2016) employed a reflexive comparison[4] approach to examine the impact of BWCs on complaints against the police and the processing of domestic violence cases in a precinct of the

[4]Reflexive evaluation or comparison compares the outcomes of the same group before and after program participation.

Phoenix, AZ, Police Department. The latter outcome could also be considered an efficiency measure. The post-test results for the officers using cameras indicated that cases were more likely to be initiated by the prosecutor compared to pre-test data (40.9% vs. 34.3%). Morrow et al.(2016) concluded that BWCs could also help improve officers' productivity in addition to reducing civilian complaints.

Finally, Ellis and colleagues (2015) assessed the effects of BWCs in Isle of Wight, UK, on a range of crime offenses, changes in criminal justice processing, complaints against officers, and officers' views on the use of BWCs. Since all police officers were issued BWCs, the results of the study also used a reflexive comparison approach. The findings related to criminal justice processes on domestic abuse cases suggest an increase in the number of cases from 3 to 21, and in 10 out the 21 cases, there was recorded footage. Furthermore, seven of these 10 cases led to arrests, and four of the seven cases led to criminal charges. The authors acknowledge that, because all officers received BWCs, the evaluation did not lend itself to any type of randomization within that police organization. Thus, in the absence of an RCT, their objective was to assess the effectiveness of BWCs from an operational angle for agencies that decide to have an agency-wide rollout of BWCs (Ellis et al., 2015).

Of the above five studies, it is worth noting the significant differences in the methodological approaches and related findings. Except for Kim's study, the two studies that used more robust evaluation approaches (i.e., RCTs) show null results or limited results compared to those that rely on a reflexive comparison approach, showing significant improvements related to the use of BWCs.

Although informative, reflexive comparison studies have serious limitations impact because these approaches attempt to examine program impacts by comparing outcomes before the intervention and after the intervention. The time difference between these two periods is considered the program's impact. These approaches generally assume that program participants' outcome would have been the same as before the intervention. Research shows, however, that this is not the case (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2016), thus limiting the validity of their findings. It is noteworthy that although these studies are limited in their statistical validity and their impacts, they still provide informative lessons concerning the implementation and operationalization of BWCs.

Despite the rapid growth in evaluations undertaken on BWCs effects on a wide range of outcomes (Lum et al., 2019), research on the effects of BWCs on police efficiency remains nascent. The strand of research presented above, which closely approximates the analysis of efficiency, offers interesting insights on the potential impacts that using BWCs might have in improving police performance. As noted above, however, none of these studies estimate a proper measure of police performance by considering how police inputs contribute to police output production. This study seeks to bridge this gap by studying how BWCs contribute, if at all, to improving police performance related to an important police output–clearance rates. Section 3 discusses the data and methods used in this research.

3 Data

The dataset constructed for 2016 consisted on information from local police agencies, crime data, and socioeconomic and demographic indicators from a variety of sources. Data on BWCs availability and use and police inputs come from the Law Enforcement Management Survey (LEMAS) (BJS, 2016). The LEMAS survey collects data from various law enforcement agencies in the US, including sheriff, state, and local agencies. In this study, the sample is limited to local police agencies because they are the law enforcement entities closest to citizens and, therefore, is where most interactions occur. Limiting the sample to local police agencies yielded efficiency estimates that more closely estimated their performance from an efficiency perspective.

Data on police outputs come from Kaplan's crime dataset (Kaplan, 2020), which contains multi-year concatenated UCR data for state and local police agencies across the US, and the socioeconomic and demographic measures come from the American Community Survey (ACS). To capture socioecononomic and demographic changes in municipalities[5], the author used ACS 5-year average data.

[5]The level of disaggregation in the ACS survey collects information on socioeconomic and demographic conditions that could affect police output production. Using a five-year average is to account for any variation in socioeconomic and demographic factors since these measures may suffer little variation from year to year.

Before discussing the empirical approach, it is worth noting that there are potential drawbacks in using non-parametric efficiency models which warrant careful consideration because biased efficiency estimates can result. Another potential limitation is the presence of missing data. Because police agency data come from the LEMAS survey, there is missing information since not all agencies respond to all questions or did not have sufficient information to answer the particular questions. Thus, to mitigate the effects of missing data on the efficiency estimates, the author eliminated those agencies from the sample where there was missing information on police inputs prior to merging it with the UCR and ACS datasets

A second potential drawback is that all non-parametric models require meeting the positivity property; that is, that all inputs and outputs must be positive numbers (>0). If this positivity property is not met, it could render values infeasible, resulting in invalid estimates because there is no possible solution to the linear programming model (Bowlin, 1998).

Research and methodological literature identify various ways to deal with this problem in DEA. One approach is to eliminate those observations with zeroes, and another is to add a sufficiently large constant so the observation meets the positivity property. This approach, while methodologically simple, can lead to an additional problem: translation invariance.

Translation invariance occurs when the addition of a constant alters the efficiency frontier and yields biased estimates since not all DEA models are translation invariant (Lovell & Pastor, 1995). It is noteworthy that Ali and Seiford (1990) developed a model that relaxes the positivity requirement by adding a constant, resulting in an affine displacement of the efficiency frontier, but does not alter it. In other words, adding a constant would push the efficiency frontier further to the right, but would not alter the frontier and, therefore, , not bias the efficiency estimates. This condition, however, works only if a constant is added to the outputs in variable returns to scale models, and to the inputs and outputs in additive modelsdescribe additive model in a footnote (Ali & Seiford, 1990; Lovell & Pastor, 1995).

As is described in the methodology section, the used of a variable returns to scale model, enabled me to fulfill the positivity and translation invariance properties by deleting those inputs with values =0 and adding a large enough constant to the outputs. Thus, after pre-processing the data to correct the potential drawbacks described above, the final study sample is comprised of 615 local police agencies.

To estimate the efficiency scores, the author followed previous literature on police efficiency and employed a model with four inputs and two outputs (Alda, 2014; Alda, Giménez, & Prior, 2019; Barros, 2007; García-Sánchez, Rodríguez-Domínguez, & Parra-Domínguez, 2013; Gorman & Ruggiero, 2008). The inputs include both the number of full-time sworn officers and non-sworn personnel, as well as number of marked and unmarked vehicles (see Table 1).

Defining police agencies' output is especially challenging as the "bottom line" of policing continues to expand and, as a result, so does its production technology[6] (Moore and Braga, 2003). The challenge is finding output measures that can capture–to the greatest extent feasible[7]–key functions of police agencies.

One measure commonly used as an output in police efficiency studies is the clearance rate (see Barros, 2007; Alda 2014; Alda et al., 2019). Clearance rates capture critical functions of police operations, such as the effectiveness of patrols, speed of police response, and police investigative capacities (Moore and Braga, 2003). In this study, the author approximated police output production by using the total number of index violent and index property crimes[8] cleared by each agency.

Index crimes are a collection of four violent and property crimes that the Federal Bureau of Investigation (FBI) uses to produce their annual crime index. Violent index crimes comprise murder, rape, robbery, and aggravated assault, and property index crimes consist of burglary, theft, motor vehicle theft, and arson[9].

Table 1 presents the summary statistics of the raw data on police inputs and outputs. On average, agencies had 266 sworn officers and 71.3 civilians (non-sworn officers); approximately 106 marked vehicles and 72 unmarked vehicles. The total number of index property crimes cleared is about twice the total number of index

[6] This refers to what police agencies do.

[7] Limitations in data availability compound this challenge.

[8] Efficiency models operate better when using units instead of rates.

[9] For an explanation of these crimes, please visit the FBI. Jacob Kaplan offers useful guidance on the advantages and disadvantages of using index crimes $vis-á-vis$ using these crimes separately.

violent crimes cleared with 560.8 and 299, respectively.

Table 1: Descriptive Statistics-Input/Output Set

Variable	Obs	Mean	Std. Dev.	Min	Max
Inputs					
Number of Sworn Officers	615	266.15	763.64	5	12042
Number of Non-Sworn Officers	615	71.28	180.26	1	2871
Number of Marked Vehicles	615	105.80	218.71	2	3797
Number of Unmarked Vehicles	615	72.19	140.78	1	1624
Outputs					
Total Index Violent Crime Cleared	615	299.99	795.74	1	12806
Total Index Property Crime Cleared	615	561.82	870.51	1	8291

Source: Own Analysis based on data from BJS(2015) and Kaplan (2020).

4 Methodological Approach

4.1 Conceptual Issues

As indicated above, modern policing has an ever-expanding ”bottom line”(Moore and Braga, 2003), and consequently, it is challenging to capture the police production function in a single model.

Production efficiency theory posits that if a decision management unit–that is, a police agency in this study–produces the same or higher output levels using the same or fewer inputs, it would be efficient relative to its peers with similar characteristics (Ray, Kumbhakar, & Dua, 2015).

A key aspect of using BWCs is that it enables officers both to resolve cases faster and reduce paperwork and, as a result, there is an increase in the number of crimes cleared (Chapman, 2018). In turn, a higher percentage of crimes cleared would lead to higher police output production. At the same time, if police increase their output production using fewer inputs (i.e., police officers) because BWCs yield more readily available data and information in the investigative process, efficiency would then improve. Furthermore, research indicates that using BWCs also helps officers increase arrests, thus leading to a quicker resolution of cases (Katz et al., 2014).

While trying to pinpoint how BWCs contribute to improving police efficiency is challenging, using an output measure, such as clearance rates, which encompasses critical police operational activities, is useful in shedding light on this issue.

4.2 Analytical Strategy

In this section, the author presents and discusses the two-staged empirical approach employed to measure the effect of BWCs on police efficiency. In the first stage, I estimated police efficiency scores using an output oriented model with variable returns to scale. In the second stage, the author used a range of matching estimators to assess the effect on police efficiency of agencies that acquired BWCs ("acquirers") compared with those agencies that did not acquire BWCs.

Matching helps balance confounding (observable) characteristics between police agencies. This approach, however, assumes that the deployment of BWCs is exogenous to police efficiency, given a set of observable characteristics. That is, if the exogeneity assumption holds, then the estimates are unbiased (Cavatassi, González-Flores, Winters, Andrade-Piedra, Espinosa, & Thiele, 2011). However, as noted earlier, it is virtually impossible to match police agencies on all the characteristics driving the adoption of BWCs. Therefore, it is possible that differences in unobservable characteristics between both groups of agencies exist and could lead to biased estimates.

To address this potential bias, the author used instrumental variable regression to examine the effect of BWCs "deployers" as compared to "acquirers" on police efficiency and thus reduce potential biases due to unobservable differences between each group. This is discussed in more detail in section 4.4.3.

4.3 First Stage

4.3.1 Data Envelopment Analysis

In the first stage, the author employed a well-known non-parametric efficiency measurement approach, Data Envelopment Analysis (DEA), to estimate the technical efficiency of local police agencies. DEA is a powerful tool for estimating organizational efficiency and has several distinct advantages as compared to, for example, parametric approaches like regression analysis.

First, DEA is flexible and can accommodate multiple inputs and outputs simultaneously; this results in obtaining a more accurate measure of efficiency of complex public-sector organizations, like the police. Second, non-parametric techniques provide information on how DMUs can improve their efficiency based on the distance from the best practice efficiency frontier. For example, the results of an output-oriented model can indicate to researchers how much output an agency could increase in order to improve efficiency (relative to the best performers) while keeping the input set constant. Third, DEA does not result in common statistical problems, like multicollinearity or heteroskedasticity, and DEA does not require normality in their distribution (Charnes, Cooper, Lewin, & Seiford, 2013), or the imposition of an '*a priori*' functional form as it is the case in regression-based models.

To estimate efficiency, DEA uses the linear combination of DMU's[10] that employ a set of inputs that are under the control of police managers–officers, vehicles– and a set of outputs that the agencies produce, that is clearance rates and crime prevented, for example.

This linear programming combination generates a "best practice" frontier, which captures the organization/s production of maximum output/s given their set inputs relative to their peers in the sample (Charnes et al., 1978). Therefore, a DMU that is on the "best practice"[11] frontier has a value of 1, indicating that, relative to its peers, it has produced more output using the same or fewer inputs and is,

[10]The DMU (Decision Management Unit) is the unit of analysis. In the case of the current study is local police agencies.

[11]This is the efficiency frontier.

therefore, more efficient.

When using frontier methodologies like DEA or similar linear programming models, defining the type of model orientation is important. There are two main models–input and output orientation. An input-oriented model measures how much a unit (for example, a police agency) could reduce its inputs while maintaining the same output level. In contrast, an output-oriented model measures how much a unit could maximize its output production with the same number of inputs. Therefore, this study employs a DEA output-oriented model with variable returns to scale (VRS).

The use of an output orientation is primarily the result of the type of output that defines police agencies' production technology. As discussed above, on of police agencies' key objective is to call offenders to "account", which is typically measured by the clearance rate (Moore & Braga, 2003 p.38). Therefore, from the point of view of police production, clearance rates are outputs the police should maximize.

The choice of variable returns to scale is also straightforward since an additional input would not result in a proportional change of the output, as it is the case of constant returns to scale models. This is because police forces normally generally operate in a non-market environments with imperfect competition and budgetary constraints (Jacobs, Smith & Street, 2006; Giménez, Keith & Prior, 2019). Consequently, police agencies often operate at an inefficient scale size. In order to support (or reject) the choice of returns to scale, the author conducted a non-parametric returns to scale test (Simar & Wilson, 2002), and the results rejected the null hypothesis (p<.01) that police agencies operate at an efficient scale[12], and thus, the choice of variable returns to scale model is appropriate.

Equation 1 below presents the basic output-oriented DEA model with variable returns to scale.

$$\text{Max } \theta$$

$s.t.$
$$\sum_{j=1}^{n} \lambda_j x_{ij} = x_{io} \qquad i = 1, 2, ..., m;$$
$$\sum_{j=1}^{n} \lambda_j y_{rj} = \theta y_{ro} \qquad r = 1, 2, ..., s;$$
$$\sum_{j=1}^{n} \lambda_j = 1 \qquad j = 1, 2, ..., n.$$

[12] This would mean that a constant returns to scale model would be more appropriate to analyze efficiency.

$\lambda_j \geqslant 0 \qquad\qquad j = 1, 2, ..., n.$

(2)

where DMU_o represents a DMU under analysis, and x_{io} and y_{ro} are the i_{th} input and r_{th} output for DMU_o.The value of θ ranges from 0 (inefficient) to 1 (efficient) . Thus, in an output-oriented model, a value of 1-θ indicates the proportional radial expansion in output that a DMU could achieve given their input set.

Despite their power and flexibility, non-parametric efficiency methods also have significant suffer from limitation limitations. For example, Because of their non-parametric nature, it renders them sensitive to the presence of outliers and measurement errors, which can lead to biased estimates. As discussed above, since the data used in this study are from a survey, the data are likely to suffer from measurement error.

Furthermore, differences in the size, location, and output produced by the agency will make some agencies outliers[13] compared to the rest of the sample because they perform significantly better than their peers. Therefore, this group of outlier agencies could define the "best practice" efficiency frontier and bias the efficiency scores downward because no other agency can perform better than this group of outliers.

Partial frontier models, such as order-m, enhance efficiency analyses and mitigate some of the common statistical problems in non-parametric techniques like DEA (Cazals, Florens, & Simar, 2002; Simar & Wilson, 2008).

Partial frontier methods operate as follows. To estimate the efficiency score, the order-m algorithm finds an m number of units (police agencies) with similar characteristics in their input/output set and then calculates how much an agency could produce using the same or fewer inputs than its peers. In this particular methodological approach, the choice of m is relevant when estimating the efficiency scores (Felder & Tauchmann, 2013). For example, choosing a value of m that is too small would yield a large share of super-efficient observations and, as the value of m increases ($m \to \infty$), the share of super-efficient observations decreases to zero[14].

[13] In efficiency analyses, outliers are also known as super-efficient or super-performers.

[14] The maximum efficiency score would be 1.

While there is not a recommended value of m, research suggests choosing a value that would yield 10% of the observations being super-efficient (Bonaccorsi, Daraio & Simar, 2006). For this study, the author chose a value of $m = 80$[15], which is considered a large value. In multi-output studies like this, however, the values of m tend to be larger than for single output studies (Felder & Tauchmann, 2013).

Furthermore, the choice of m enables the detection of outliers, which helps explain whether there are particular characteristics of these agencies that make them outliers as compared to their peers (Daraio & Simar, 2007). In addition, because the efficiency frontier is not bounded from above at 1, outperforming agencies (outliers) can yield efficiency scores that are larger than 1 and will not bias efficiency estimates downward. Consequently, the resulting efficiency estimates are closer to the 'true' efficiency frontier compared to a DEA model (Daraio & Simar, 2007). This last feature is potentially useful in studying police forces because their inherent heterogeneity will be reflected in internal organizations, practices, use of resources, and, ultimately, in the production process itself. When performing efficiency analyses of police forces, outliers will emerge, and this technique enables researchers to understand why those observations in the sample perform significantly better than their peers.

4.4 Second Stage-Matching and Instrumental Variable Regression

4.4.1 Matching

In the second stage of this study, the author employed a range of matching estimators to assess the effect of BWCs on police efficiency. Because this study is based on survey and administrative data, it was not possible to randomly assign agencies into treatment and control groups. Therefore, to be able to compare the effects of agencies that acquired and deployed BWCs with those that did not, it was necessary to create groups that were roughly similar based on a set of observable characteristics.

[15]The author conducted efficiency analyses for different values of m. They are not reported here but available upon request.

Matching methods enable researchers to generate a credible counterfactual–that is, what would the efficiency levels be in the absence of BWCs?–, by creating two comparable groups based on observable characteristics. Consequently, the results on the efficiency scores could be attributed to the effect of having integrated BWCs into their policing functions. In addition to being able to generate comparable groups, matching methods reduce selection bias (Cavatassi et al., 2011; Guo Fraser, 2010).

To match police agencies, I used two questions from LEMAS survey:

- Has your agency acquired body-worn cameras?
- Have body-worn cameras been deployed to officers in your agency?

The first question enabled the construction an Intent to Treat (ITT) variable comprising all the agencies that acquired BWCs (”acquirers”) regardless of whether or not they deployed them. The second question enabled the construction of a Treatment on the Treated (TOT) measure which captures all agencies that had acquired BWCs and deployed them with their officers (”deployers”). It was possible to generate the latter measure because, according to the responses of the survey, 84% of agencies that acquired BWCs had implemented a partial or full deployment.

Some researchers have argued that some studies of BWCs suffer from potential selection effects because agencies choose to adopt BWCs technology for a variety reasons, such as consent decrees, the agency's interest in improving their performance, accountability and legitimacy, state and local laws mandates, or organizational characteristics (Maskaly, Donner, Jennings, Ariel, & Sutherland, 2017). For example, larger police agencies may have budgets to adopt and fully implement this type of technology. Nowacki and Willits (2016), however, showed that this is not always the case. In their study of organizational drivers of adoption of BWCs, they found that agencies prone to using technology in their operational activities appear more likely to adopt innovative technologies such as BWCs. Conversely, the size of operational budgets and the presence of unions appear to hinder the adoption of this type of technology in order to prevent limitations in police discretion.

Matching methods can also help reduce potential selection biases associated with the adoption of BWCs as well as minimizing Type I errors (Guo & Fraser, 2014). As discussed above, however, it requires both a strong exogeneity assumption and no lurking unobservable variables that could bias the results. Because police agencies are complex organizations (Maguire, 2003), it is virtually impossible to match agencies on all the variables influencing the adoption of BWCs. The author tried to address this issue by first matching agencies on a set of exogenous factors and internal organizational characteristics influencing the adoption of BWCs, and then conducting additional tests to determine whether the results could be affected by hidden bias due to the influence of unobservable characteristics. In the next sections, the author presents and explains the data, the empirical strategy, and the findings.

Although matching algorithms can yield consistent and robust estimates on the effects of BWCs on police efficiency, using only the ITT sample would yield conservative results because the ITT sample includes those agencies that only acquired BWCs and those agencies that deployed BWCs with their officers. (Gupta, 2011). This would underestimate the effect of the actual deployment of BWCs somewhat because agencies may acquire BWCs, but might be non-compliant due to limited capacity and organizational management to effectively deploy BWCs (Hyland, 2018; Nowacki & Willits, 2016). The author employed an instrumental variable (IV) regression to conduct the TOT analysis to examine the effect of BWCs "deployers" compared to "non-deployers" and obtain a more precise estimate of the effect of BWCs on police efficiency.

The TOT analysis yields what is known as the Local Average Treatment Effects (LATE) estimates. Imbens and Angrist (1994) argue that LATE estimates capture the average treatment effect among those exposed to the treatment. In this study, it would capture the effects on the efficiency of those police agencies that deployed BWCs.

4.4.2 Matching Estimators

Matching can be accomplished in several ways. One of the most well-known methods is propensity score matching (PSM). The propensity score defined as the probability of receiving treatment conditional on a set of observable baseline characteristics $ei = Pr(Z_i = 1|X_i)$ (Rosenbaum & Rubin, 1983) (see 2 for the variables used to match agencies).

To estimate the propensity, the author used a probit regression model[16] to predict the probability of being treated by an intervention. The propensity score allowed the author to construct two similar groups[17] based on a set of observable covariates. Thus, any differences in the levels of efficiency between these groups can be attributed to the adoption of BWCs.

To examine causal effects using observational data, Rosenbaum and Rubin(1983) argued that two assumptions must be met. The first assumption is the ”unconfoundedness assumption”,which states that outcomes on the treatment and control groups are independent of participation status conditional on a set of observable covariates (X). This is illustrated with the following equation:

$$(Y(0), Y(1)) \perp\!\!\!\perp D|X$$

The second assumption that must be met in propensity score matching is the ”overlap assumption”, which states that observations with the same observable values can be in the treatment or control group (Caliendo & Kopeinig, 2008). The following equation illustrates the overlap condition:

$$0 < P(D = 1|X) < 1$$

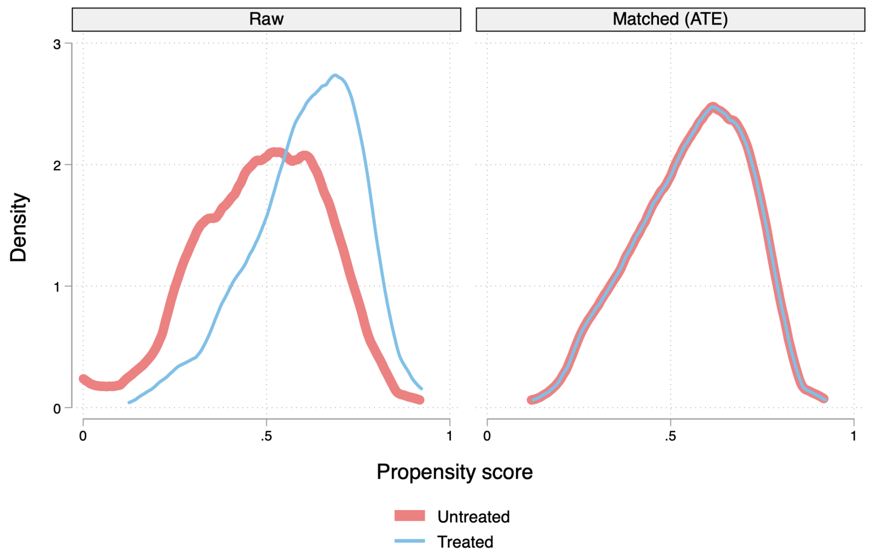
Figure 1 illustrates the density curves before and after matching using the propensity score. After matching, the figure shows no significant differences between BWCs ”acquirers” and ”non-acquirers”.

Recently, however, matching methods such as PSM have sparked a debate about its effectiveness in generating balanced samples to assess impact. For example, King and Nielsen (2019) argue that PSM might achieve the opposite of a balanced sample leading to inefficiency, model dependence, and biased estimates (King &

[16]Probit and logistic regression models are the most common approaches to estimating the propensity score, although researchers have examined other approaches.

[17]To remind the readers, the two groups the author created are: ”Acquirers” and ”Non-Acquirers” (ITT) and ”Deployers” and ”Non-Deployers” (LATE)

Figure 1: Density Curves-Unmatched vs. Matched Samples



Nielsen, 2019 p.2; Iacus, King & Porro, 2012). To address these shortcomings, the authors proposed a new approach–Coarsened Exact Matching (CEM). This approach finds exact matches, one with that has adopted BWCs and one that has not, instead of matching on propensity scores.

The CEM approach coarsens the exogenous covariates, divides them into different strata, and then performs an exact matching within each stratum (King & Nielsen, 2019). A major trade-off of matching is that it requires researchers to choose which covariates to match agencies with. This challenge is evident when using CEM because if the strata are too complex, the likelihood of finding an exact match is lowee and, it is more difficult to conduct estimations (Vigneri & Lombardini, 2017). Jann (2017), however, argues that it is possible to conduct matching when using algorithms that do not throw away good matches[18]. In this study, the author used a wide range of matching algorithms, including CEM, in order to check the consistency of the results across various models.

[18]Jann (2017) argues that the results presented by King and colleagues appear to be based on the worst possible matching approach: one to one exact matching without replacement.

4.4.3 Instrumental Variable Regression

While useful and informative, the ITT analyses may not provide an accurate estimate of BWCs effects on police efficiency since matching methods rely on the assumption that the adoption of BWCs is exogenous to the outcome given a set of observable characteristics X_i as shown in equation (1) above. The main advantage of using an IV approach, when a valid instrument can be found, is that it deals with potential bias from observable and unobservable differences in BWCs adopters and non-adopters. This method can also be used to test the exogeneity assumption used in propensity score matching (Ravallion, 2005). Relaxing the exogeneity assumption, however, requires finding a valid instrument, which must be strongly correlated with the adoption of BWCs but it cannot be correlated with the error term. In impact evaluation studies, it is common to use ITT as an instrument since, as evident in this study, all police agencies that acquired BWCs have the option of deploying them, but not every agency does. As noted earlier, out of 84% of in the sample that deployed BWCs, only 40% deployed them fully with their officers.

To estimate the LATE effects, the IV approach requires two stages, and each stage is illustrated in the equations below:

$$\text{BWCs}_i = \delta Z_i + \varphi X_i + v_i \qquad \text{(First Stage)}$$
$$\theta_i = \beta X_i + \widehat{BWCs}_i + \epsilon_i \qquad \text{(Second Stage)}$$

where the first stage captures the relationship between instrument Z_i and the adoption of BWCs, and φ captures the relationship between instrument X_i and the adoption of BWCs. In the second stage of the 2SLS model, $\widehat{BWCs}_i$ captures the predicted adoption of BWCs estimated in the first stage. The variables v_i and ϵ_i are the error terms of the first and second stage of the model (Cavataassi et al., 2011).

The first stage is estimated as a linear probability model. Angrist (2000) suggests using this approach when the first stage is a limited dependent variable model and argues that it is consistent and safer since using other models, such as probit/logit, in the first stage will only be consistent if the model is exactly correct.

The author used two measures of BWCs deployment to conduct the IV analyses.

The first variable captured those agencies that implemented a partial deployment of BWCs with their officers, and, the second variable captured those agencies that permanently deployed BWCs with their officers. The estimates on the full BWCs deployment are likely to be larger than the partial deployment because agencies that partially deployed BWCs did it for testing purposes or for particular assignments and, consequently, may not exploit the benefits of this technology.

Table 2 present summary statistics of the set of observable characteristics used to match police agencies and as explanatory variables in the IV regressions. Based on prior research and theoretical tenets in organizational theory, the author used a set of exogenous and organizational characteristics that could influence the adoption of BWCs (Alda, 2017; Alda & Dammert, 2019; Alda, Giménez, & Prior, 2019; Barros, 2007; Gorman & Ruggiero, 2008). These include total population, population density, the unemployment rate, the GINI coefficient of income inequality, the poverty rate, the adoption of other technology, the number of prevented civilian complaints against officers; and important organizational structure characteristics, such as the size of the police agency, operational budget; and measures of organizational complexity, such as functional and vertical differentiation. The first of the variables of organizational complexity captures how a police agency assigns tasks within its organization, and it is measured by the number of specialized units in each agency (Nowacki & Willits, 2018; Maguire, 2003). The second organizational complexity variable measures the hierarchy within an agency, and it is measured by the midpoint salary difference between the highest and lowest rank officer (Nowacki & Willits, 2018; Maguire 2003).

Table 2: Summary Statistics-Observable Characteristics

Variable	Obs	Mean	Std. Dev.	Min	Max
Population Density	615	3738.05	4450.13	217.56	53766.98
Population Estimate (2012)	615	124195.46	259902.74	702	3857799
% Population with less than High School	615	13.46	7.68	0	55.17
Unemployment Rate	615	7.86	3.16	1	22.15
GINI Coefficient of Income Inequality	615	0.45	0.047	0.31	0.62
Poverty Rate	615	16.24	7.74	0.61	43.25
All crimes recorded	615	6645.55	14905.18	0	172294
Civilian Complaints (Reciprocal*)	495	0.17	0.262	0.001	1
Police Agency Size	615	2.62	0.53	1	3
Acquired Car Dashboard Cameras	600	0.71	0.454	0	1
Budget (Ln)	598	16.47	1.417	12.723	20.951
Functional Differentiation	602	4.652	6.690	0	137
Vertical Differentiation	581	69116.98	34697.86	1195	246771

Source: BJS (2015), Kaplan (2020).

* The reciprocal value approximates the total number of civilian complaints prevented by each agency.

5 Results

Table 3 presents the overall efficiency estimates and the estimates disaggregated by police size. The mean efficiency score was 0.76, indicating that, on average, police agencies that are inefficient relative to the best performers could increase their outputs (crimes cleared) by 31 percent (from 0.76 to 1). Larger and smaller police agencies perform better with efficiency scores between 0.84 and 0.79, respectively. The efficiency score for mid-size police agencies was 0.60, suggesting that they performed worse compared with their larger and smaller peers.

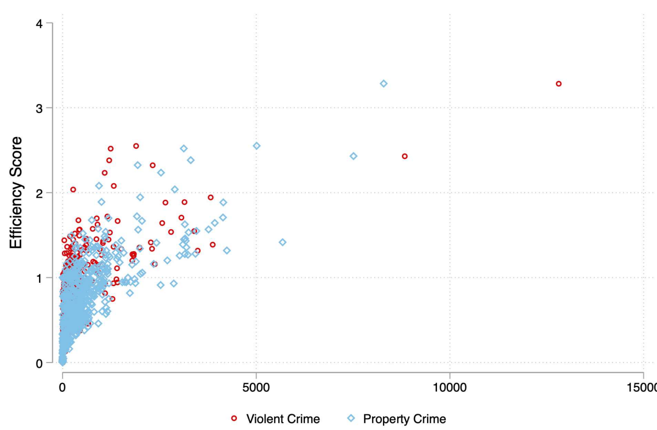
Table 3: Order-m Efficiency Estimates

	Mean	Std. Dev.	Min	Max
Overall Efficiency Score	0.76	0.45	0.00	3.28
Police agency (1-10 Officers)	0.79	0.30	0.25	1.00
Police agency (11-100 Officers)	0.60	0.40	0.01	1.68
Police agency (¿100 Officers)	0.84	0.46	0.00	3.28

Source: Own Analyses using BJS (2015), Kaplan (2020).

Figure 2 illustrates the efficiency results by output. It reflects the maximum level of output produced by municipal police forces given their inputs. Police forces with values at or above 1 indicate that they performed better in their output production than the number of m agencies used as comparators.

Figure 2: Order-m Scores



There is, however, significant variation in the levels of efficiency. Out of 615 agencies, only 28 were efficient ($\theta = 1$), less than 5% of the sample and were distributed between small and mid-size agencies. It is worth noting that some agencies were very inefficient and others were super-efficient relative to their peers, with efficiency scores as high as 3.3. To explain this result in more detail, an agency with an efficiency score of 3.3 means that it cleared as much as three times more output than a similar m number of peers. In the Annex, figure 3 presents the same results without outlier agencies–$\theta > 1$– and more clearly shows the variation in police performance.

Table 4 presents the estimates on the effects of BWCs on police efficiency using a range of matching estimators and instrumental variable regression. The ITT results that agencies that acquired BWCs have a positive, strong, and statistically significant effect on police efficiency. The estimates are remarkably robust and consistent across model specifications. Improvements in efficiency range from eight to 12 percentage points, depending on the model. The regression adjustment model yielded the smallest coefficient, whereas the mahalanobis distance estimator yielded the largest coefficient. Regression adjusted models in matching estimators add an additional layer of robustness because they reduce additional bias in the covariate balance, ensuring consistency in the estimates, which might explain a slightly smaller estimate in the analyses (Abadie & Imbens, 2011 p.1).

In regard to the IV estimates, the first stage criteria show that the ITT is a valid instrument in the model because it is positive, strong, and statistically significant in the first stage and the instrumented variable is also positive, strong, and highly significant in the second stage. The F-statistic rejects the null hypothesis that the instrument is weak with values well over the accepted 'rule of thumb' threshold of $F > 10$[19] (Cuesta & Alda, 2012). Furthermore, tests for over-identification and endogeneity assumptions show that there are no over-identifying restrictions and the tests accept the null hypothesis that the instrument can be treated as exogenous. The latter supports the exogeneity assumption needed for the matching estimators (Cavatassi et al., 2011).

As expected, the IV (LATE) estimates are larger in magnitude than the ITT

[19] New research questions the use of the $F > 10$ as the rule of thumb for first stage estimates. Lee and colleagues (2020) suggest that F-statistic values should be larger than 104.7 in order to have a true 5% t-ratio test. As Table 4 shows, the first stage F-statistic value is >104.7

estimates. This is because the LATE estimates capture the effect of BWCs on those agencies that deployed BWCs compared to those agencies that acquired BWCs but did not deploy them. The results indicate that agencies that deployed and permanently deployed BWCs improved their efficiency between 12 and 21 percentage points, respectively. This suggests that controlling for both observable and unobservable characteristics, agencies that deployed BWCs experienced greater efficiency gains, thus supporting the argument that the use of BWCs can help improve police efficiency.

Table 4: Regression Results

	DM[1]	PS[2]	RM[3]	NN-3[4]	NN-5[5]	RA-MD[6]	DWPS[7]	CEM[8]	IV[9]	IV-2[10]
Efficiency	0.124***	0.100**	0.103**	0.112**	0.105**	0.079**	0.086**	0.109**	0.125***	0.209***
	(0.033)	(0.042)	(0.040)	(0.044)	(0.044)	(0.038)	(0.037)	(0.040)	(0.043)	(0.072)
Constant								0.669***	-0.427	-0.687
								(0.0314)	(0.503)	(0.522)
Observations	446	446	446	446	446	446	446	415	446	446
R^2								0.02	0.30	0.30

[1] MD = Malahanobis Distance Matching.
[2] PS = Propensity Score Matching.
[3] RM = Propensity Score Ridge Matching.
[4] NN-3 = Nearest Neighbor Matching (3).
[5] NN-5 = Nearest Neighbor Matching (5).
[6] RA-MD = Regression Adjustment.
[7] DWPS = Doubly Weighted Propensity Score Matching.
[8] CEM = Coarsened Exact Matching.

[9] 2SLS Instrumental Variable Regression. First stage F-statistic, 301.3, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 11.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates..

[10] 2SLS Instrumental Variable Regression-2 First stage F-statistic, 268.11, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates..

Notes:
All matching methods except for CEM were done using Stata's user-written command kmatch (Jann, 2019). The CEM analyses were done using Stata's user-written command CEM (King, 2019). Standard Errors in Parenthesis: significance *10%, **0.05%, ***0.01%

Table 5 presents the predicted efficiency scores for each matching algorithm and the IV models for each group of police agencies; that is, ”acquirers” vs. ”non-acquirers” and ”acquirers” vs. ”deployers”[20]. The predicted efficiency scores are significantly higher, about ten percentage points in the ITT analyses and 20 percentage points larger between ”acquirers” and ”deployers” in the LATE results.

Table 5: Predicted Efficiency Scores

	Non-Acquirers	Acquirers	Acquirers	Deployers*,**
MD	0.678	0.802		
PS	0.694	0.795		
RM	0.693	0.796		
NN-3	0.706	0.818		
NN-5	0.712	0.817		
RA	0.712	0.798		
DWPS	0.718	0.798		
CEM	0.667	0.783		
IV			0.667	0.859
IV-2			0.667	0.860
Avg.	0.695	0.797	0.667	0.859

* Partial Deployment. ** Full Deployment.

5.1 Robustness Checks

Although the results are consistent across matching and IV specifications, the presence of outliers could drive the second stage estimates, given that the proportion of super-efficient agencies is relatively large. Therefore, to check whether these outliers drive the second stage results, the author dropped from the sample those agencies with efficiency scores larger than one and re-estimated the matching and IV models. Table 6 in the Annex presents the result and show, on average, slightly smaller effects, although the LATE estimates are slightly larger than those of the preferred models in Table 4 above.

As an additional robustness test, the author re-estimated the efficiency scores using the reduced sample; that is, the resulting sample after eliminating the observations

[20]Predicted efficiency scores for deployers include those agencies that partially deployed BWCs and agencies that implemented a full deployment of BWCs.

that had efficiency scores > 1[21]. Table 6 in the Annex presents the estimates. The results are still strong and statistically significant across matching estimators and the IV regressions, and do not substantially alter the results of the preferred model specifications (see Table 4 above). The average of all the effects are slightly larger in the preferred model specifications–0.115 vs. 0.109 percentage points–, which is driven by the ITT estimates.

The reduced sample of the original order-m scores to $\leqslant 1$ shows that BWCs improve efficiency between eight and 11 percentage points for the ITT estimates and 13 to 23 percentage points for the LATE estimates. Conversely, the results on the re-estimated efficiency scores on the reduced sample (see Table 7 in Annex) also show positive, strong, statistically significant effects of BWCs on police efficiency. The magnitude of the coefficients ranges from 13 to 16 percentage points for the ITT estimates and from 20 to 34 percentage points for the LATE estimates. The coefficients are larger probably as a result of the sample being reduced by 166 agencies. Also, the efficiency estimates have changed because the number and type of comparators (agencies) in the sample differ from the base sample and that invariably influences the generation of the efficiency frontier.

The author also conducted the matching and IV analyses on the group of super-efficient police agencies ($\theta > 1$) (see Table 8 in the Annex). These results indicate no effects of BWCs on efficiency among the super-performing agencies[22].

A concern with efficiency estimation is the potential imbalance in the data because of differing magnitudes in inputs and outputs. One way to address this issue in DEA and DEA-based analyses is to mean-normalize the data to ensure similarity in inputs and outputs across units (Sarkis, 2007). The author proceeded to mean-normalize the inputs and outputs, estimate the efficiency scores, and use them as the outcome in the matching and IV analyses.

The results indicate that the mean efficiency scores were slightly lower than the preferred model–0.68 compared to 0.76 (see Table 10 in the Annex). This reduction in the efficiency scores is likely the result of mean-normalizing the data, which may

[21]The reader should note that even after dropping outlier observations, the analyses will still yield super-efficient observations.

[22]It is worth noting that the N for these analyses is substantially smaller–166– and will likely affect the results.

lessen the influence of outlier agencies in the model. The matching and IV results are smaller in magnitude compared to the preferred models, but are still positive and statistically significant (see Table 9 in the Annex). For the ITT analyses, the effects of BWCs on police efficiency range from six to ten percentage points, and for the IV models, effects range from 11 to 18 percentage points.

Finally, Table 12 in the Annex presents the estimates of a basic DEA model using an output-oriented and variable returns to scale model. As discussed above it is plausible that super-efficient agencies may drive the efficiency scores.Therefore, an order-m model would prevent these agencies from setting the efficiency frontier–at = 1 and introduce bias by pushing the rest of the units downward and causing a higher percentage of agencies to become inefficient (Epstein & Henderson, 1989).

The results show a significant drop in efficiency scores to an average score of 0.46 compared to the average of 0.76 in the order-m model. These results help validate the use of an order-m model to obtain more accurate efficiency estimates. Table [redacted] presents the matching and IV estimates. Similar to the preferred models, the ITT results indicate that acquiring BWCs has a positive and statistically significant effect on police efficiency. The ITT estimates range from four to seven percentage points[23]. Similarly, the IV estimates are positive and statistically significant, and the size of the coefficients indicate effects ranging from five to 10 percentage points.

5.2 Hidden Bias

The author further checked the sensitivity of the results to the presence of hidden bias driven by unobservable factors that could influence the adoption of BWCs. As noted earlier, several internal and external organizational factors and operational factors can influence decision-making in the adoption of BWCs. Therefore, the results should not rule out the possibility of the presence of hidden bias. Gangl and DiPrete (2004) argue that although propensity score matching[24] removes most of the bias due to observable characteristics, it is not a consistent estimator if hidden

[23] The regression adjustment estimates are positive but no longer statistically significant at conventional levels ($p < .05$).

[24] Note that propensity score matching is one of several matching algorithms the author used in the analyses.

bias is present (DiPrete and Gangl, 2004, p.272).

5.2.1 Rosenbaum Bounds

First, the author used the Rosenbaum bounds test to examine how the results would be affected in the presence of hidden bias from an unobserved confounding variable. It is noteworthy that the presence of hidden bias does not invalidate the results rather, they convey important information on how large the effect of an unobserved variable must be in order to change the conclusions inferred from the original estimates (DiPrete & Gangl, 2004).

To conduct the analysis, the author set the maximum value for Γ, at 1 with increments of 0.1, which are considered appropriate for these type of data (Keele, 2010). Γ values start at 1 and indicate no presence of unobserved confounders, and the p-value should hold if there is no hidden bias present. The results suggest that the critical value Γ at which the p-value is no longer statistically significant at conventional values is equal to 1.7 (see Table 15 in the Annex). Thus, in order to question the study's results, an unobserved variable would have to affect the log odds of adoption of BWCs by a factor of 1.7.

5.2.2 Simulated Confounder

Second, the author used the simulated confounder approach proposed by Ichino, Mealli, and Nannicini (2008). This approach assumes that a binary variable $\bigcup$ can be simulated and used as another observable characteristic in the matching analysis. The primary underlying assumption of this approach is that the both the observable characteristics and the simulated confounder can influence the adoption of BWCs.

The results demonstrate the extent to which the baseline estimates are robust to the failure of the conditional independence assumption. The author employed two variables to conduct the simulated confounder analyses on the original outcome variable–police efficiency. The first variable is the size of the police[25], and the

[25]Generating the simulated confounder requires a binary variable. Thus, I generated one where

second variable is the use of dashboard computers. Both variables are likely associated with the adoption of BWCs. Using a nearest neighbor and kernel matching. Table 16 in the Annex presents the results showing positive and statistically significant effects of both the baseline and the simulated confounder model. The coefficient of 0.11 suggests negligible differences between the baseline and the simulated confounder estimates. Furthermore, as recommended in Ichino et al. (2008), both the outcome and selection effects are positive (>1). Like the Rosenbaum bounding approach, these results confirm the robustness of the estimates in the preferred models.

5.2.3 Relative Correlation Restrictions

Finally, the author used the relative correlation restrictions (RCR) methodology proposed by Krauth (2016) to construct informative bounds on the effects of BWCs on police efficiency and assess how these estimates behave to deviations from the exogeneity assumptions (Krauth, 2016, p. 2). This methodology assumes a correlation between the adoption of BWCs and the unobserved variables relative to the correlation between the variable of interest and the observed exogenous characteristics. The author examined the potential effect of a correlation between the adoption of BWCs and unobservable characteristics that is 0.25, 0.5, 0.75, 1, and twice the correlation size between the adoption of BWCs and the observable characteristics the author employed for the matching and IV analyses (Desai & Joshi, 2013).

Table 17 in the Annex presents the results. The first row shows the OLS regression point estimates in the absence of hidden bias (λ=0), while the remaining rows present the point estimates for up to twice the correlation between the adoption of BWCs and observable characteristics. The RCR results suggest that the point estimates are robust to a weak correlation–0 and 10 percent– between the adoption of BWCs and observable characteristics. Although the bounds on the effect are narrow and close to the OLS estimate, however, these are not statistically significant at conventional levels. Furthermore, the RCR bounds show no effect at moderate or large correlations ($0 \leq \lambda \leq 1$) as the bounds include 0. Thus, the RCR

large police agencies take a value of 1 and 0 otherwise.

results may raise concern on the influence of unobserved confounding variables on the matching estimates.

Overall, the signs and magnitudes of the effects of BWCs on police efficiency are robust to different matching estimators and potential hidden bias.

6 Conclusions and Limitations

This study examined the effect of BWCs on police efficiency using a sample of local police agencies in the U.S. in 2016. The author conceptualized the adoption of BWCs across local police agencies as those agencies that acquired BWCs and those that deployed them, either partially or fully, with their officers. This differentiation created two groups: (1) an Intent to Treat (ITT) group for all the agencies that acquired BWCs, and (2) a Treatment on the Treated (TOT) group for those agencies deploying them. To examine the effects of BWCs on police efficiency, the author employed a two-stage analytical approach.

In the first stage, the author estimated the levels of police efficiency using an efficiency model that is robust to the presence of outliers and measurement error inherent in administrative and survey data. Specifically, the author used an output-oriented and variable returns to scale model because organizations like the police should maximize the output produced (clearance rates) using the same or fewer inputs.

In regard to efficiency, the estimates suggest that police agencies have room for improvement. The efficiency scores range from 0.60 to 0.84, depending on the police agency's size, with an overall mean of 0.76. In other words, on average, police agencies could improve their performance by increasing 31 percent of their output production–that is, their clearance rates– using the same or fewer inputs. Furthermore, the results showed that over 100 agencies were deemed super-efficient. This means that these agencies produced between more than 1 (efficiency score >1), and as much as three times more output than similar peers using the same number of inputs.

In the second stage of the analyses, the author employed a range of matching estimators and instrumental variable analyses using the efficiency scores as the outcome of interest. The results were positive, strong, and statistically significant across all matching and IV models. The ITT estimates suggest an improvement in efficiency between seven and 12 percentage points, and the LATE estimates suggest an improvement in efficiency ranging from ten to about 21 percentage points. The effects on efficiency gains are substantial. For example, if police can increase efficiency by an average of 11 percentage points[26],the number of crimes cleared would increase from an average of 430 violent and property crimes cleared to an average of 494. Such an increase amounts to an average of 64 more crimes cleared annually through the deployment of BWCs.

The author also conducted robustness tests and examined the sensitivity of the results to the presence of hidden bias. The robustness tests suggested that, after re-analyzing the models, the presence of outliers does not affect the estimates' strength and robustness, and, if anything, the magnitude of the effects increased from an average of 11 to 12 percentage points. The sensitivity analyses suggest that the models are robustness to the presence of hidden bias except for the relative correlation restrictions approach. The RCR results showed mild robustness to the presence of unobserved factors that could question the robustness of the estimates in the preferred models. Altogether, the findings of this study provide strong support to the argument that the adoption of BWCs can contribute to improving police efficiency, among other aspects of policing.

There are several important caveats to keep in mind with this study. First, the study sample is limited to only local police agencies. The LEMAS survey collects data on a much larger sample of law enforcement agencies and includes the sheriff, county, and state police, among others. Therefore, any inferences based on these results should be attributed to local police agencies and not as effects that can be generalized across law enforcement agencies. In addition, due to data limitations and missing data for a number of agencies, the data required pre-processing and, as a result, ended up reducing the sample to 615 local police agencies.

Second, there are limitations related to the number and types of police inputs. The LEMAS survey does not contain data on key inputs in a police production function, such as computers, phones, and GPS, among others. The use of

[26]This is the average of all the regression coefficients in the LATE models in Table 4.

technology, paired with adequate organizational and management changes, is important in improving efficiency (Garicano & Heaton, 2010; Milgrom & Roberts, 1990). For this study's purposes, the author was able to use two key police inputs, which are the number of police officers and the number of civilian personnel.

Third, the author could not capture in the analyses the variation in the adoption of BWCs. The data indicate that some agencies had acquired BWCs 10-15 years ago, and some acquired them as recently as 2016, the year the BWCs survey was implemented. Since 2012 the number of police agencies that have adopted BWCs increased by more than 500% from 19 in 2013 to 121 in 2015[27] (see Figure 3 in the Annex). Therefore, it is possible that the early adoption of BWCs may have influenced the efficiency results since those agencies had more time to become familiar with using this technology. One possible way to address this issue in future research is to conduct temporal analysis and estimate yearly efficiency levels since the shape of the efficiency frontier, and the units that generate it may change from year to year.

Finally, although this study deliberately focused on local police agencies, they still face variations in their technology sets due to differences in organizational structure, financial and human resources, and the operating environment. For example, the efficiency results indicate that the number of super-efficient agencies is somewhat large and driven by mid-size and large agencies. While the methods used in the first and second stages helped address differences between agencies significantly[28], variation still exists in agencies' technology sets, which ultimately affects the generation of the efficiency frontier (O'Donnell, Rao, & Battese, 2008). Thus, modeling the production frontier to account for differences in technology sets would yield efficiency estimates that compared the performance of agencies with peers having similar technology sets. Unfortunately, sample size limitations did not permit the author to model police production function under different technology sets.

Considering these caveats, the findings nevertheless raise an important question on the mechanisms through which the use of BWCs improve police efficiency. This is important from an operational point of view. It is challenging to shed light a priori

[27]This is based on this study's sample.

[28]Note that eliminating the super-efficient observations did not substantially alter the estimates.

on how BWCs cameras could improve police efficiency, given limitations in data that would allow researchers to model the complexity of a police agency's production function. This study, however, offers some potential channels.

Research shows that using BWCs generally contributes to reducing the time needed to clear a crime and send it to the next phase within the criminal justice system (c.f. Morrow et al., 2016). Furthermore, historical research on clearance rates appears to provide support to this argument. Scott and colleagues (2019) suggest average historical trends, despite showing significant stability, there was substantial variation among agencies in their clearance rate performance. Organizational changes and other factors were the primary drivers of variation (Scott, Wellford, Lum & Vovak, 2019).

Another potential channel is the compounding effect that BWCs can have on improved performance as a result of faster police response times. For example, recent evidence suggests that faster police response times can improve crime clearance rates by as much as 4.7% (Vidal & Kirchmeier, 2018). If faster response times alone can lead to higher clearance rates, the enhanced data and information that BWCs can collect could be an influential factor in improving clearance rates.

Of course, organizational factors and external factors beyond police managers' control invariably influence an agency's performance (Alda & Dammert, 2019). As Scott and colleagues(2019) suggested, differences in organizational characteristics could explain variation in clearance rate performance. Hence, having adequate organizational factors conducive to a full deployment of BWCs, and training on proper use of BWCs and other available technology, can positively impact efficiency (Milgrom & Roberts, 1990). Ultimately, however, officers must be compliant in using and exploiting this technology's capabilities to improve law enforcement practices, particularly around maximizing output production while using the same or fewer resources.

Improving police organizations' efficiency can significantly impact budgetary allocations in local government and police organizations to ensure proper allocation of resources to maximize service delivery. Taken together, the results of this study shed light on the effects that this technology has on police efficiency. Expanding on this strand of research will become increasingly important in the growing body of literature on the use of BWCs by law enforcement agencies.

References

Abadie, A., & Imbens, G. W. (2011). Bias-corrected matching estimators for average treatment effects. Journal of Business Economic Statistics, 29(1), 1-11.

Alda, E. (2014). ”How are Police Doing in Combating Crime? An Efficiency Analysis of the Policía Nacional Civil in Guatemala.” Policing: An International Journal of Police Strategies and Management 37(1): 87-107.

Alda, E. & Dammert, L. (2019). ”Weathering the Storm!. The Effects of the External Environment on Police Performance in Perú”. Policing: An International Journal of Police Strategies and Management.

Ali, I., & Seiford, L. M. (1990). Translation invariance in data envelopment analysis. Operations Research Letters, 9(6), 403–405.

Barros, C. P. (2007). The city and the police force: analysing relative efficiency in city police precincts with data envelopment analysis. International Journal of Police Science Management, 9(2), 164-182.

Blanes i Vidal, J., & Kirchmaier, T. (2018). The effect of police response time on crime clearance rates. The Review of Economic Studies, 85(2), 855-891.

Bogetoft, P., & Otto, L. (2010). Benchmarking with DEA, SFA, and R (Vol. 157). Springer Science & Business Media.

Bonaccorsi, A., Daraio, C., & Simar, L. (2006). Advanced indicators of productivity of universities: An application of robust nonparametric methods to Italian data. Scientometrics, 66(2), 389-410.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. Journal of economic surveys, 22(1), 31-72.

Cavatassi, R., González-Flores, M., Winters, P., Andrade-Piedra, J., Espinosa, P., & Thiele, G. (2011). Linking smallholders to the new agricultural economy: The case of the plataformas de concertación in Ecuador. Journal of Development Studies, 47(10), 1545-1573.

Chapman, B. (2018). Body-worn cameras: What the evidence tells us. National Institute of Justice, December. Retrieved from:https://nij.ojp.gov/topics/articles/body-worn-cameras-what-evidence-tells-us

Charnes, A., & Cooper, W. W. (1957). Management models and industrial applications of linear programming. Management science, 4(1), 38-91.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2(6), 429-444.

Cuesta, J. & Alda, E. (2012). The effects of trust on victimization in Colombia. Journal of Peace Research, 49(6), 833-846.

Desai, R. M., & Joshi, S. (2013). Can producer associations improve rural livelihoods? Evidence from farmer centres in India. Journal of Development Studies, 50(1), 64-80.

DiPrete, T. A., & Gangl, M. (2004). 7. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. Sociological methodology, 34(1), 271-310.

Ellis, T., Jenkins, C., & Smith, P. (2015). Evaluation of the introduction of personal issue body worn video cameras (Operation Hyperion) on the Isle of Wight: Final report to Hampshire Constabulary.

Epstein, M. K., & Henderson, J. C. (1989). Data envelopment analysis for managerial control and diagnosis. Decision Sciences, 20(1), 90-119.

Felder, S.,& Tauchmann, H. (2013). Federal state differentials in the efficiency of health production in Germany: an artifact of spatial dependence?. The European Journal of Health Economics, 14(1), 21-39.

García-Sánchez, I. M., Rodríguez-Domínguez, L., & Parra-Domínguez, J. (2013). Yearly evolution of police efficiency in Spain and explanatory factors. Central European Journal of Operations Research, 21(1), 31-62.

Garicano, L., & Heaton, P. (2010). Information technology, organization, and productivity in the public sector: Evidence from police departments. Journal of Labor Economics, 28(1), 167-201.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). Impact evaluation in practice. The World Bank.
https://www.bjs.gov/content/pub/pdf/BWCslea16.pdf

Gorman, M. F., & Ruggiero, J. (2008). Evaluating US state police performance using data envelopment analysis. International Journal of Production Economics, 113(2), 1031-1037.

Guo, S., & Fraser, M. W. (2014). Propensity score analysis: Statistical methods and applications (Vol. 11). SAGE publications.

Harcourt, B. E., & Ludwig, J. (2006). Broken windows: New evidence from New York City and a five-city social experiment. The University of Chicago Law Review, 271-320.

Hyland, S. S. (2018). Body-worn cameras in law enforcement agencies, 2016. US Department of Justice, Bureau of Justice Statistics, Washington. https://www.bjs.gov/content/pub/pdf/BWCslea16.pdf

Iacus, S. M., King, G., & Porro, G. (2012). CEM: software for coarsened exact matching. https://gking.harvard.edu/files/cem.pdf

Ichino, A., Mealli, F., & Nannicini, T. (2008). From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity?. Journal of applied econometrics, 23(3), 305-327.

Imbens, G. (2010). Methods for Estimating Treatment Effects IV: Instrumental Variables and Local Average Treatment Effects. Technical report, Lecture Notes 2, Local Average Treatment Effects, Impact Evaluation Network, Miami. http://www.cedlas-er.org/sites/default/files/cer_ien_activity_files/miami_late.pdf

Jann, B. (2017). Why propensity scores should be used for matching. 2017 German Stata Users Group Meeting Berlin, June 23, 2017 https://www.stata.com/meeting/germany17/slides/Germany17_Jann.pdf.

Jennings, W. G., Fridell, L. A., Lynch, M., Jetelina, K. K., & Reingle Gonzalez, J. M. (2017). A quasi-experimental evaluation of the effects of police body-worn cameras (BWC) on response-to-resistance in a large metropolitan police department. Deviant behavior, 38(11), 1332-1339.

Kaplan, J.(2020). Jacob Kaplan's Concatenated Files: Uniform Crime Reporting (UCR) Program Data: Hate Crime Data 1991-2018. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], https://doi.org/10.3886/E103500V6.

Katz, C. M., Choate, D. E., Ready, J. R., & Nuño, L. (2014). Evaluating the impact of officer worn body cameras in the Phoenix police department. Phoenix, AZ: Center for Violence Prevention Community Safety, Arizona State University. Retrieved from: publicservice.asu.edu/sites/default/files/ppd_spi_feb_20_2015_final.pdf

King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. Political Analysis, 27(4), 435-454.

Kim, T. (2019). The Impact of Body Worn Cameras on Police Use of Force and Performance (October 23, 2019). Available at SSRN: https://ssrn.com/abstract=3474634

Krauth, B. (2016). Bounding a linear causal effect using relative correlation restrictions. Journal of Econometric Methods, 5(1), 117-141.

Lee, D. S, McCrary, J., Moreira, M. J., & Porter, J. (2020). Valid t-ratio Inference for IV. Retrieved from: https://arxiv.org/abs/2010.05058v1

Lovell, C. K., & Pastor, J. T. (1995). Units invariant and translation invariant DEA models. Operations research letters, 18(3), 147-151.

Lum, C., Koper, C. S., Merola, L., Scherer, A., & Reioux, A. (2015). Existing and ongoing body worn camera research: Knowledge gaps and opportunities (Report for the Laura and John Arnold Foundation). Fairfax, VA: Center for Evidence-Based Crime Policy, George Mason University. Retrieved from: https://cebcp.org/wp-content/technology/BodyWornCameraResearch.pdf

Maguire, E. R. (2003). Organizational structure in American police agencies: Context, complexity, and control. SUNY Press. Albany, NY.

Maskaly, J., Donner, C., Jennings, W. G., Ariel, B., & Sutherland, A. (2017). The effects of body-worn cameras (BWC) on police and citizen outcomes. Policing: An International Journal of Police Strategies Management.

Milgrom, P., & Roberts, J. (1990). The economics of modern manufacturing: Technology, strategy, and organization. The American Economic Review, 511-528.

Moore, M. H., & Braga, A. (2003). The "bottom line" of policing: What citizens should value (and measure!) in police performance. In Police Executive Research Forum.

Nowacki, J. S., & Willits, D. (2018). Adoption of body cameras by United States police agencies: an organisational analysis. Policing and society, 28(7), 841-853.

O'Donnell, C. J., Rao, D. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical economics, 34(2), 231-255.

Ravallion, M. (2005) Evaluating anti-poverty programs. World Bank Policy

Research Working Paper no. 3625, Washington, DC.

Ray, S. C., Kumbhakar, S. C., & Dua, P. (2015). Benchmarking for performance evaluation. Springer India 2015. New Dehli, India. https://doi.org/10.1007/978-81-322-2253-8

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Sarkis, J. (2007). Preparing your data for DEA. In "Modeling data irregularities and structural complexities in data envelopment analysis" (pp. 305-320). Springer, Boston, MA.

Scott, T. L., Wellford, C., Lum, C., & Vovak, H. (2019). Variability of crime clearance among police agencies. Police Quarterly, 22(1), 82-111.

Simar, L., & Wilson, P. W. (2002). Non-parametric tests of returns to scale. European Journal of Operational Research, 139(1), 115-132.

Tone, K., & Sahoo, B. K. (2004). Degree of scale economies and congestion: A unified DEA approach. European Journal of Operational Research, 158(3), 755-772.

Tunca, H., & Yesilyurt, F. (2016). Hospital efficiency in Turkey: metafrontier analysis. European Scientific Journal, ESJ, 12(10).

Trumbull, W., Witte, A. (1981). Determinants of the costs of operating large-scale prisons with implications for the costs of correctional standards. Law & Society Review, 16(1), 115-137.

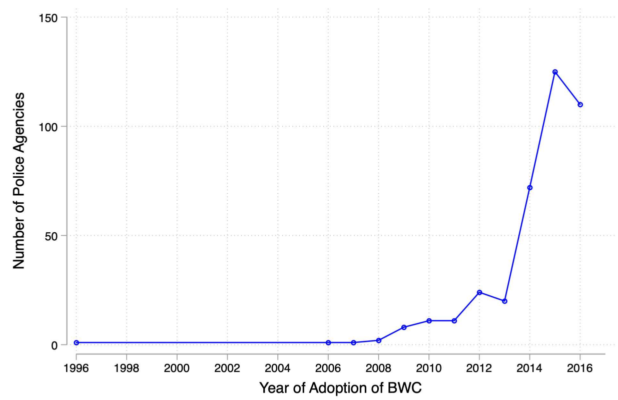
BJS-United States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. Law Enforcement Management and Administrative Statistics (LEMAS), 2015. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2015-09-22. https://doi.org/10.3886/ICPSR36164.v2

U.S. Census Bureau. (2017). 2012-2016 American Community Survey 5-year Public Use Microdata Samples [Stata Data file]. Retrieved from https://www.socialexplorer.com/explore-tables

Vigneri, M., & Lombardini, S. (2017). Women's Empowerment in Mali: Impact evaluation of the educational project:'Girls CAN-Promoting Secondary Education in West Africa'. https://oxfamilibrary.openrepository.com/bitstream/handle/10546/620252/er-womens-empowerment-mali-effectiveness-review-100417-en.pdf;jsessionid=F4B608525A340777AB42F2342309A255?sequence=1

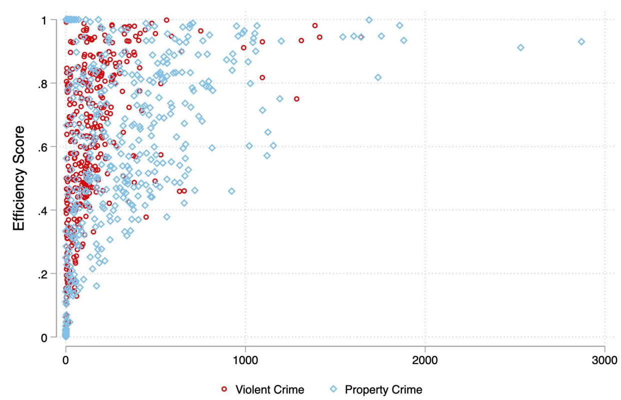
Annex-Supplementary Figures and Tables

Figure 3: Yearly Adoption of BWCs



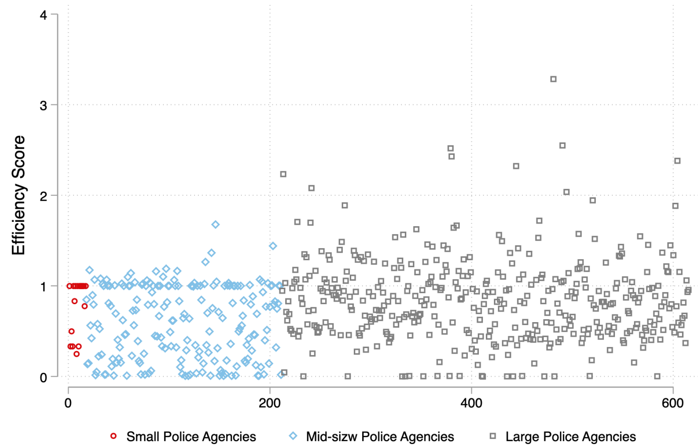
Notes: Source: Own Analyses using BJS (2015), Kaplan (2020).

Figure 4: Order-m Scores



Notes: Order-m efficiency scores without outlier agencies.

Figure 5: Efficiency by Agency Size



Notes: Order-m efficiency scores by agency size.

Figure 6: Bivariate Plot: Efficiency Scores vs. Efficiency Scores-Mean Normalized

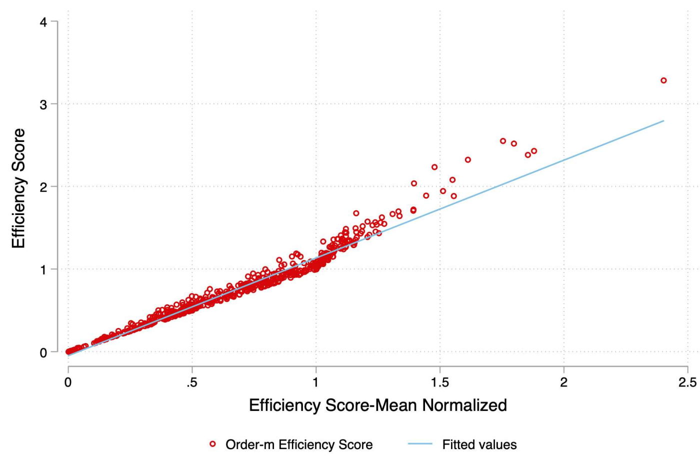


Table 6: Results-Sample with Efficiency Scores $\leqslant 1$

	DM[1]	PS[2]	RM[3]	NN-3[4]	NN-5[5]	RA-MD[6]	DWPS[7]	CEM[8]	IV[9]	IV-2[10]
Efficiency	0.111***	0.094**	0.093**	0.0790*	0.090**	0.0827*	0.096***	0.0850***	0.133***	0.230***
	(0.0339)	(0.0377)	(0.0379)	(0.0413)	(0.0397)	(0.0424)	(0.0367)	(0.0327)	(0.0390)	(0.0675)
Constant								0.517***	0.823	0.447
								(0.0252)	(0.557)	(0.582)
Observations	320	320	320	320	320	320	320	316	320	320
R^2								0.02	0.13	0.12

[1] MD = Malahanobis Distance Matching.
[2] PS = Propensity Score Matching.
[3] RM = Propensity Score Ridge Matching.
[4] NN-3 = Nearest Neighbor Matching (3).
[5] NN-5 = Nearest Neighbor Matching (5).
[6] RA-MD = Regression Adjustment.
[7] DWPS = Doubly Weighted Propensity Score Matching.
[8] CEM = Coarsened Exact Matching.
[9] 2SLS Instrumental Variable Regression. First stage F-statistic, 760.27 ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 11.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates.

[10] 2SLS Instrumental Variable Regression-2 First stage F-statistic, 162.20, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates.

Notes: All matching methods except for CEM were done using Stata's user-written command kmatch (Jann, 2019). The CEM analyses were done using Stata's user-written command CEM (King, 2019). Standard Errors in Parenthesis: significance *10%, **0.05%, ***0.01%

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
Estimates are based on 1,000 bootstrap replications.

Table 7: Results - Re-Analyses of Efficiency Scores and BWCs Effects

	DM[1]	PS[2]	RM[3]	NN-3[4]	NN-5[5]	RA-MD[6]	DWPS[7]	CEM[8]	IV[9]	IV-2[10]
Efficiency	0.167***	0.151***	0.153***	0.110***	0.048***	0.032***	0.049***	0.0.050**	0.196***	0.338***
	(0.0462)	(0.0442)	(0.0456)	(0.0512)	(0.0262)	(0.0263)	(0.0262)	(0.0259)	(0.0496)	(0.0870)
Constant								0.360***	0.302	-0.252
								(0.0204)	(0.641)	(0.677)
Observations	320	320	320	320	446	446	446	416	320	320
R^2								0.01	0.23	0.20

[1] MD = Malahanobis Distance Matching.
[2] PS = Propensity Score Matching.
[3] RM = Propensity Score Ridge Matching.
[4] NN-3 = Nearest Neighbor Matching (3).
[5] NN-5 = Nearest Neighbor Matching (5).
[6] RA-MD = Regression Adjustment.
[7] DWPS = Doubly Weighted Propensity Score Matching.
[8] CEM = Coarsened Exact Matching.
[9] 2SLS Instrumental Variable Regression. First stage F-statistic, 760.27, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 11.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates..

[10] 2SLS Instrumental Variable Regression-2 First stage F-statistic, 162.20, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates. .

Notes: All matching methods except for CEM were done using Stata's user-written command kmatch (Jann, 2019). The CEM analyses were done using Stata's user-written command CEM (King, 2019). Standard Errors in Parenthesis: significance *10%, **0.05%, ***0.01%

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
Estimates are based on 1,000 bootstrap replications.

Table 8: Results - Robustness Analyses-Outlier Agencies

	DM[1]	PS[2]	RM[3]	NN-3[4]	NN-5[5]	RA-MD[6]	DWPS[7]	CEM[8]	IV[9]	IV-2[10]
Efficiency	0.0661	0.0208	0.0145	0.0391	0.0355	0.0273	-0.0147	0.000814	0.0501	0.0764
	(0.0548)	(0.0529)	(0.0575)	(0.0517)	(0.0507)	(0.117)	(0.0521)	(0.0851)	(0.0463)	(0.0713)
Constant								1.302***	-0.933	-1.029
								(0.0721)	(0.711)	(0.738)
Observations	126	126	126	126	126	126	126	71	126	126
R^2								0.00	0.44	0.42

[1] MD = Malahanobis Distance Matching.
[2] PS = Propensity Score Matching.
[3] RM = Propensity Score Ridge Matching.
[4] NN-3 = Nearest Neighbor Matching (3).
[5] NN-5 = Nearest Neighbor Matching (5).
[6] RA-MD = Regression Adjustment.
[7] DWPS = Doubly Weighted Propensity Score Matching.
[8] CEM = Coarsened Exact Matching.
[9] 2SLS Instrumental Variable Regression. First stage F-statistic, 742.95, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates..

[10] 2SLS Instrumental Variable Regression-2 First stage F-statistic, 117.03, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates. .

Notes: All matching methods except for CEM were done using Stata's user-written command kmatch (Jann, 2019). The CEM analyses were done using Stata's user-written command CEM (King, 2019). Standard Errors in Parenthesis: significance *10%, **0.05%, ***0.01%

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
Estimates are based on 1,000 bootstrap replications.

Table 9: Results - Robustness Analyses-Normalized Input/Output Set

	DM[1]	PS[2]	RM[3]	NN-3[4]	NN-5[5]	RA-MD[6]	DWPS[7]	CEM[8]	IV[9]	IV-2[10]
Efficiency	0.103***	0.0892**	0.0919**	0.0969**	0.0916**	0.0684**	0.0763**	0.0940***	0.110***	0.184***
	(0.0355)	(0.0374)	(0.0383)	(0.0376)	(0.0364)	(0.0343)	(0.0379)	(0.0344)	(0.0379)	(0.0632)
Constant								0.608***	0.0471	0.182
								(0.0270)	(0.432)	(0.446)
Observations	446	446	446	446	446	446	446	416	446	446
R^2								0.02	0.24	0.24

[1] MD = Malahanobis Distance Matching.
[2] PS = Propensity Score Matching.
[3] RM = Propensity Score Ridge Matching.
[4] NN-3 = Nearest Neighbor Matching (3).
[5] NN-5 = Nearest Neighbor Matching (5).
[6] RA-MD = Regression Adjustment.
[7] DWPS = Doubly Weighted Propensity Score Matching.
[8] CEM = Coarsened Exact Matching.
[9] 2SLS Instrumental Variable Regression. First stage F-statistic, 1271.66, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates..

[10] 2SLS Instrumental Variable Regression-2 First stage F-statistic, 246.22, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates. .

Notes: All matching methods except for CEM were done using Stata's user-written command kmatch (Jann, 2019). The CEM analyses were done using Stata's user-written command CEM (King, 2019). Standard Errors in Parenthesis: significance *10%, **0.05%, ***0.01%

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
Estimates are based on 1,000 bootstrap replications.

Table 10: Order-m Efficiency Estimates- Mean Normalized

	Mean	Std. Dev.	Min	Max
Overall Efficiency Score	0.68	0.27	0.0003	2.40
Police agency (1-10 Officers)	0.78	0.30	0.25	1.00
Police agency (11-100 Officers)	0.57	0.38	0.006	1.16
Police agency (>100 Officers)	0.73	0.36	0.0003	2.40

Source: Own Analyses using BJS (2015), Kaplan (2020).

Table 11: Predicted Efficiency Scores- Mean Normalized

	Non-Acquirers	Acquirers	Acquirers	Deployers*,**
MD	0.618	0.721		
PS	0.625	0.714		
RM	0.623	0.715		
NN-3	0.636	0.733		
NN-5	0.641	0.732		
RA	0.646	0.714		
DWPS	0.63	0.706		
CEM	0.607	0.701		
IV				0.607 0.7619
IV-2				0.607 0.771
Avg.	0.62825	0.717 0.607 0.76645		

Source: Own Analyses using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
* Partial Deployment. ** Full Deployment.

Table 12: Results - Robustness Analyses using a DEA model

	DM[1]	PS[2]	RM[3]	NN-3[4]	NN-5[5]	RA-MD[6]	DWPS[7]	CEM[8]	IV[9]	IV-2[10]
Efficiency	0.0732***	0.0524**	0.0557**	0.0542**	0.0484*	0.0322	0.0481*	0.0501*	0.0584**	0.0975**
	(0.0251)	(0.0266)	(0.0271)	(0.0264)	(0.0262)	(0.0263)	(0.0260)	(0.0259)	(0.0283)	(0.0473)
Constant								0.360***	0.383	0.262
								(0.0203)	(0.338)	(0.348)
Observations	446	446	446	446	446	446	446	415	446	446
R^2								0.01	0.19	0.19

[1] MD = Malahanobis Distance Matching.
[2] PS = Propensity Score Matching.
[3] RM = Propensity Score Ridge Matching.
[4] NN-3 = Nearest Neighbor Matching (3).
[5] NN-5 = Nearest Neighbor Matching (5).
[6] RA-MD = Regression Adjustment.
[7] DWPS = Doubly Weighted Propensity Score Matching.
[8] CEM = Coarsened Exact Matching.
[9] 2SLS Instrumental Variable Regression. First stage F-statistic, 1271.66, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates..

[10] 2SLS Instrumental Variable Regression-2 First stage F-statistic, 246.22, ($p < .01$).
Kleibergen-Paap rank statistic for cluster-robust 2SLS (null hypothesis is that the equation is under-identified) is rejected.
Stock-Yogo critical value (at 95% confidence) for weak-instrument test statistics (Kleibergen-Paap Wald or CraggDonald F) is 16.38 for maximum bias of IV estimator to be no more than 10% of the maximal IV size (inconsistency) of OLS estimates. .

Notes: All matching methods except for CEM were done using Stata's user-written command kmatch (Jann, 2019). The CEM analyses were done using Stata's user-written command CEM (King, 2019). Standard Errors in Parenthesis: significance *10%, **0.05%, ***0.01%

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
Estimates are based on 1,000 bootstrap replications.

Table 13: DEA Efficiency Estimates

	Mean	Std. Dev.	Min	Max
Overall Efficiency Score	0.41	0.27	0.0001	1.00
Police agency (1-10 Officers)	0.65	0.37	0.05	1.00
Police agency (11-100 Officers)	0.36	0.30	0.002	1.00
Police agency (>100 Officers)	0.41	0.24	0.0001	1.00

Source: Own Analyses using BJS (2015), Kaplan (2020).

Table 14: Predicted Efficiency Scores-DEA Model

	Non-Acquirers	Acquirers	Acquirers	Deployers*,**
MD	0.361	0.434		
PS	0.368	0.420		
RM	0.366	0.422		
NN-3	0.372	0.426		
NN-5	0.377	0.425		
RA	0.389	0.421		
DWPS	0.374	0.422		
CEM	0.359	0.409		
IV			0.358	0.446
IV-2			0.358	0.447
Avg.	0.370	0.422	0.358	0.446

Source: Own Analyses using BJS (2015), Kaplan (2020), and US Census Bureau (2017).
* Partial Deployment. ** Full Deployment.

Table 15: Rosenbaum Bounds

Γ	sig^{+}	sig^{-}	t-hat^{+}	t-hat^{-}	CI^{+}	CI^{-}
1.0	0.0000	0.0000	0.1490	0.1490	0.0796	0.2120
1.1	0.0002	0.0000	0.1310	0.1664	0.0595	0.2295
1.2	0.0009	0.0000	0.1133	0.1822	0.0439	0.2451
1.3	0.0037	0.0000	0.0971	0.1961	0.0277	0.2595
1.4	0.0114	0.0000	0.0815	0.2106	0.0127	0.2724
1.5	0.0286	0.0000	0.0667	0.2221	-0.0017	0.2850
1.6	0.0602	0.0000	0.0557	0.2346	-0.0137	0.2956
1.7	0.1099	0.0000	0.0445	0.2445	-0.0252	0.3063
1.8	0.1785	0.0000	0.0342	0.2536	-0.0371	0.3147
1.9	0.2636	0.0000	0.0233	0.2631	-0.0477	0.3247
2.0	0.3599	0.0000	0.0138	0.2713	-0.0583	0.3335

Γ- Log odds of differential assignment due to unobserved factors.
sig^{+}-Upper bound significance level.
sig^{-}-Lower bound significance level.
t-hat^{+}-Upper bound Hodges-Lehmann point estimate.
t-hat^{-}-Lower bound Hodges-Lehmann point estimate.
CI^{+}-Upper bound confidence interval (a= .95).
CI^{-}-Lower bound confidence interval (a= .95).
Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).

Table 16: Simulated Confounder

Police Size	Baseline Estimate	Simulated Estimate	Outcome Effect	Selection Effect
Kernel Matching	0.114***	0.112***	1.44	1.525
Nearest Neighbor	0.057	0.106	1.576	1.538
Dashboard Cameras				
Kernel Matching	0.114***	0.114***	1.081	1.833
Nearest Neighbor	0.056	0.11	1.045	1.829

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017). Estimates are based on 1,000 bootstrap replications.

Table 17: Relative Correlation Restrictiions

	ITT	TOT
OLS point estimate ($\lambda = 0$)	0.106***	0.110***
(95% CI)	(0.03,0.178)	(0.04,0.183)
Bounds, $0 \leqslant \lambda \leqslant 0.1$	[0.112,0.260]	[0.29,0.444]
(95% CI)	(0.09,0.106)	(0.10,0.112)
Bounds, $0 \leqslant \lambda \leqslant 0.25$	[-0.212,0.260]	[-0.409, 0.444]
(95% CI)	(0.065,0.106)	(0.082,0.112)
Bounds, $0 \leqslant \lambda \leqslant 0.5$	[-0.405,0.260]	[-0.611,0.444]
(95% CI)	(0.206,0.106)	(0.050,0.112)
Bounds, $0 \leqslant \lambda \leqslant 1$	[-0.920,0.260]	[-1.00,0.444]
(95% CI)	(-0.081,0.106)	(-0202,0.112)
Bounds, $0 \leqslant \lambda \leqslant 2$	[-3.10,0.260]	[-1.883,0.444]
(95% CI)	(-0.390,0.106)	(-0.204,0.112)
λ_∞	2.82	3.34
$\lambda(0)$	0.61	2.94
Minimum λ for which bounds include zero	0.61	2.94

Source: Own analysis using BJS (2015), Kaplan (2020), and US Census Bureau (2017).

Notes: λ is the assumed correlation between the treatment and the observed variables. Bounds reflect the estimates of the adoption of BWCs (ITT and TOT) on police efficiency. Intervals in brackets are the estimated rcr bounds and the intervals in parenthesis are 95% asymptotic confidence intervals.